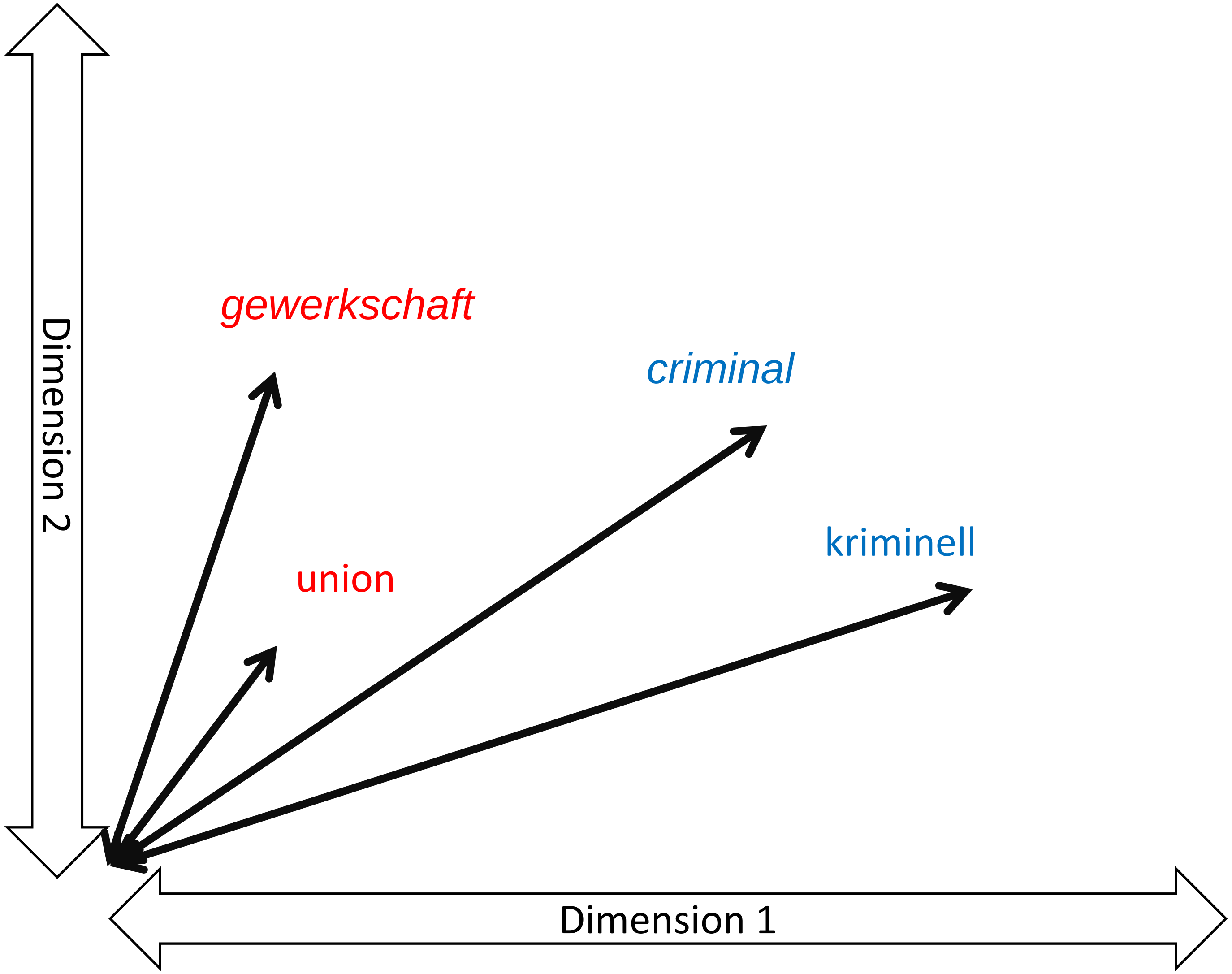
Session 5

How similar are texts and in what regard?

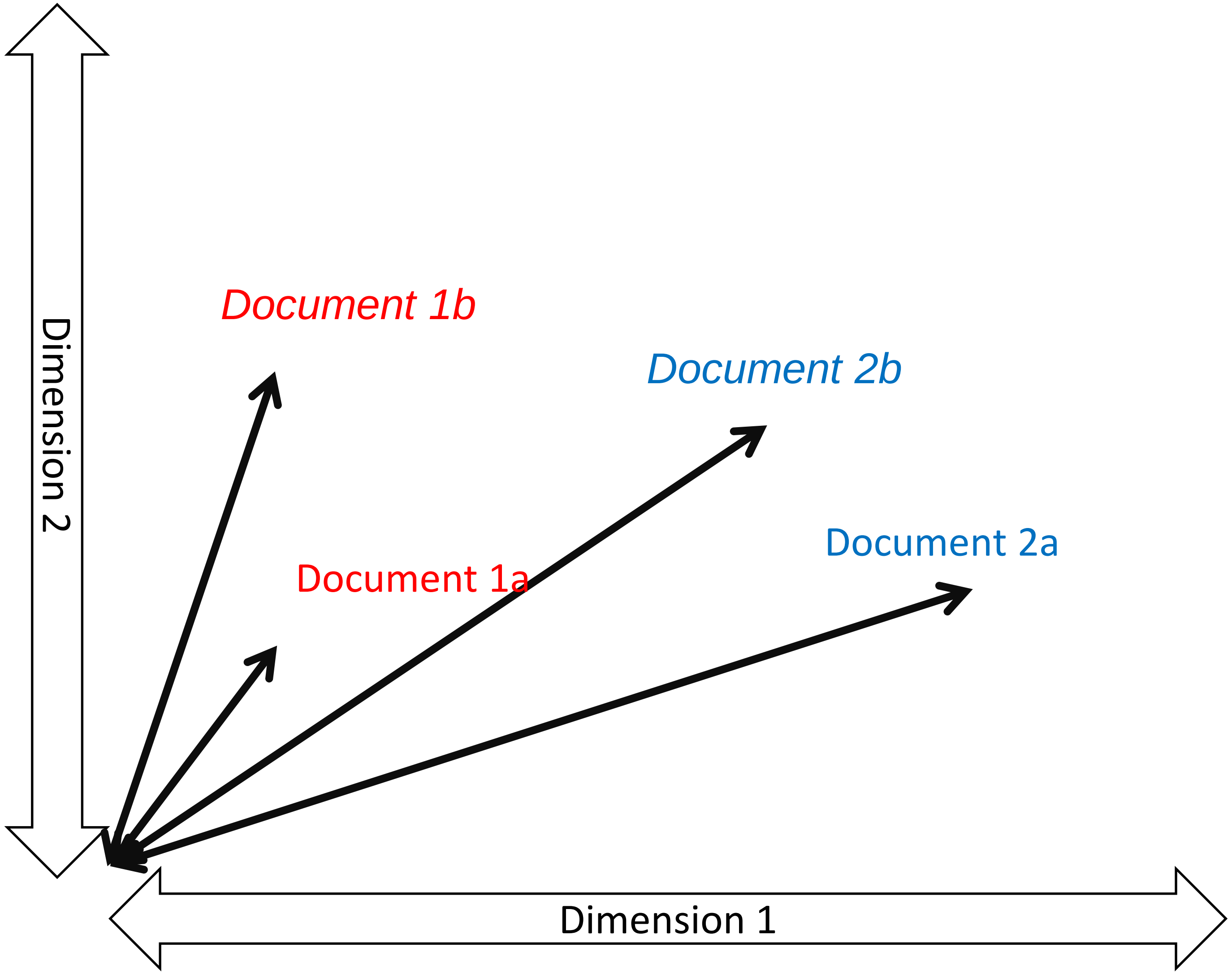
Is it the similarity in my variables?

Text Analysis as Similarity



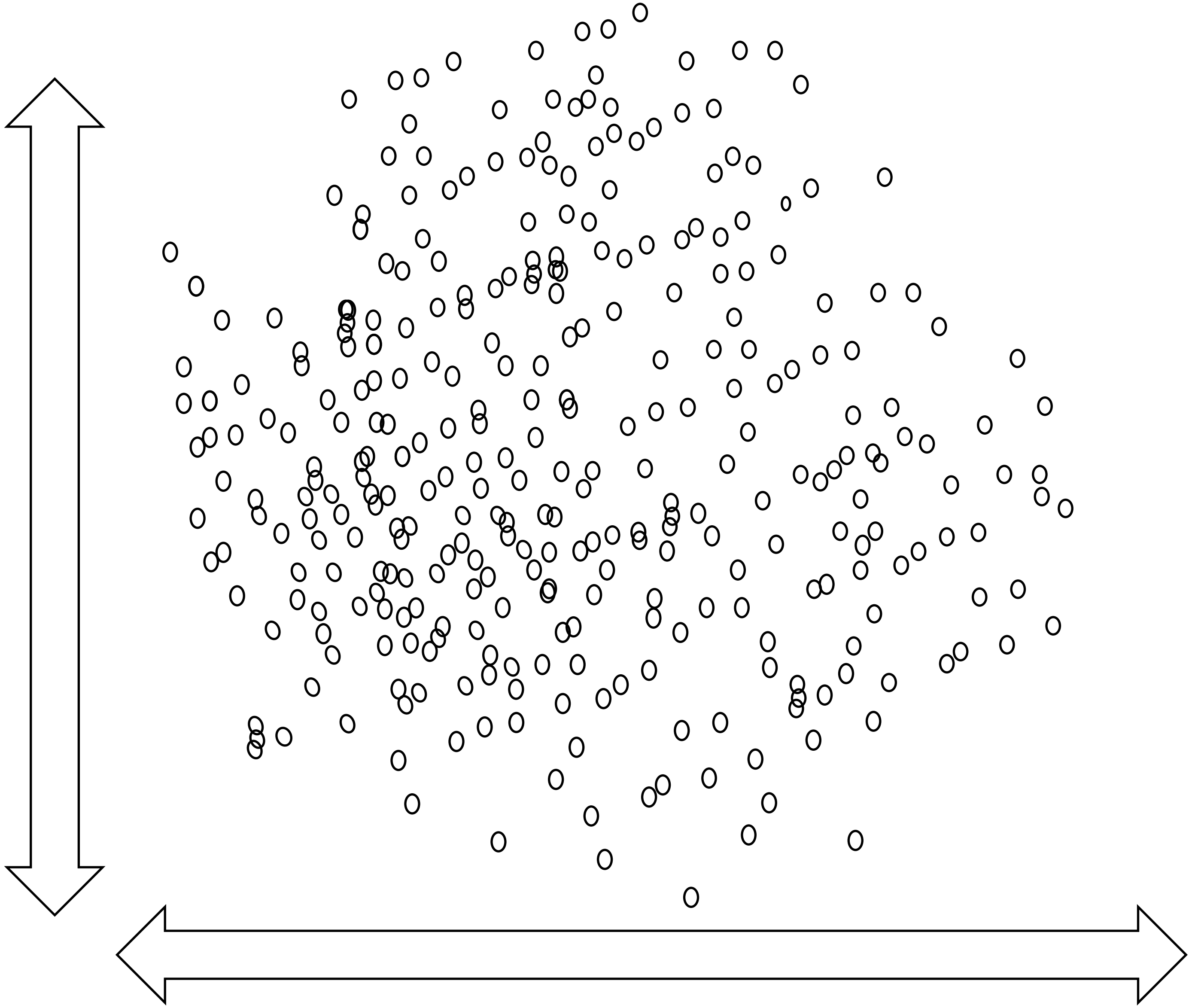
Every word is embedded in a semantic space

Text Analysis as Document Similarity

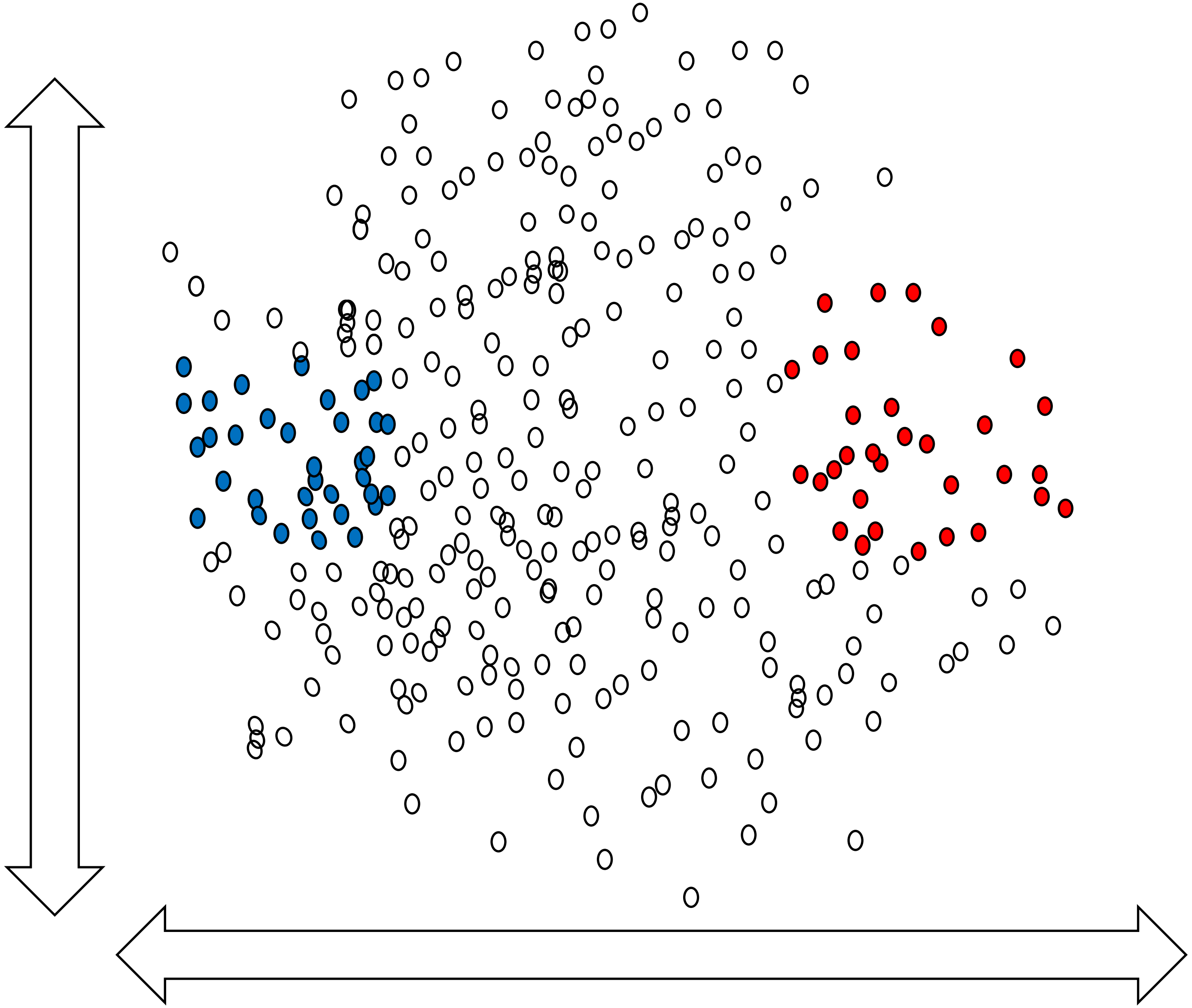


Every document is embedded in a semantic space

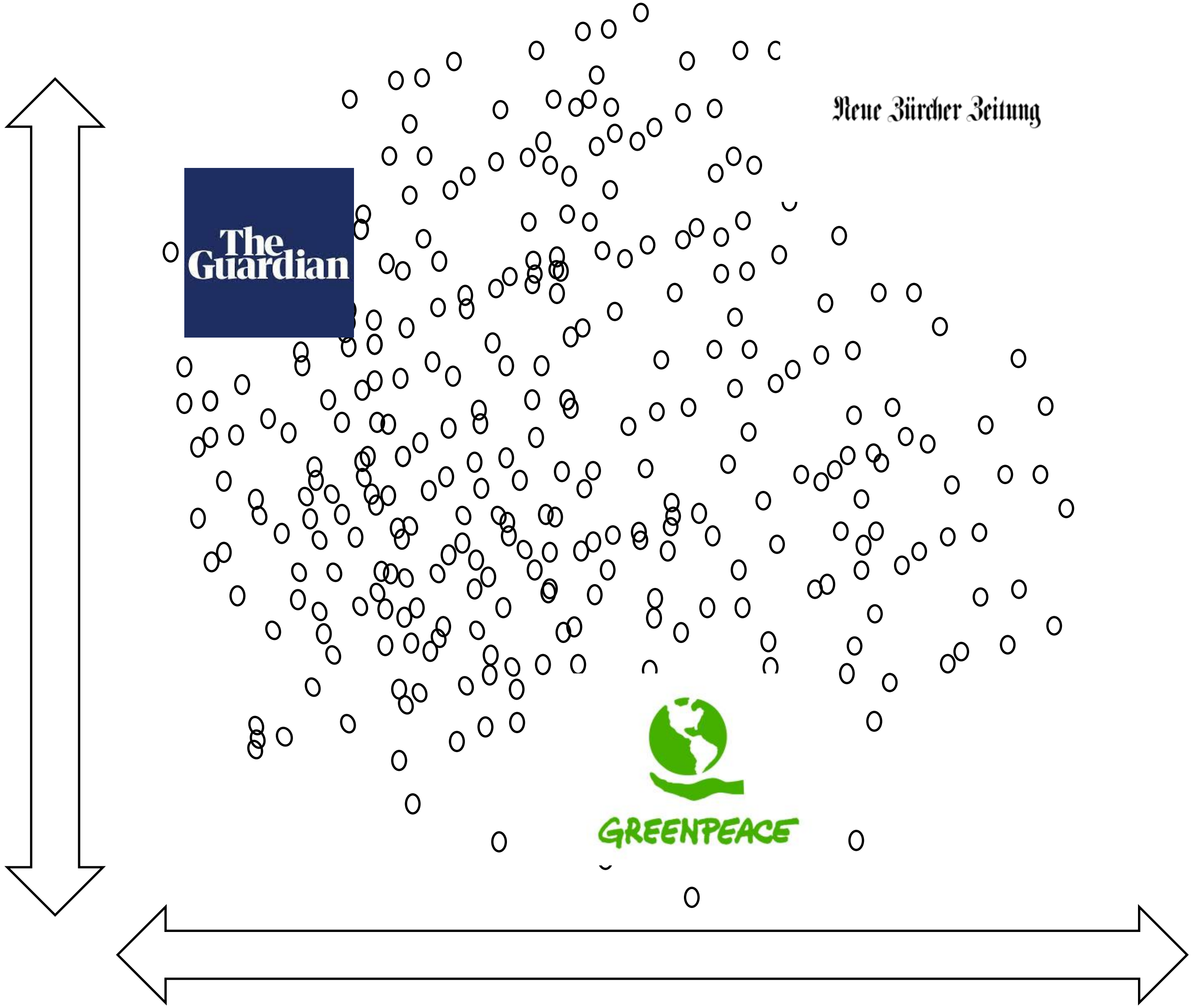
Aim of text analysis: find the dimension that corresponds to your variable of interest!



Aim of text analysis: find the dimension that corresponds to your variable of interest!



Aim of text analysis: find the dimension that corresponds to your variable of interest!



How to find the Dimension?

- Unsupervised: Hope for the best!
- Semisupervised: Give your model a hint
- Supervised: Tell it the endpoints
- Problems: Additional dimensions correlated with your variable

Every Problem can be a classification Problem

- Define Intension
- Define Extension (borders)

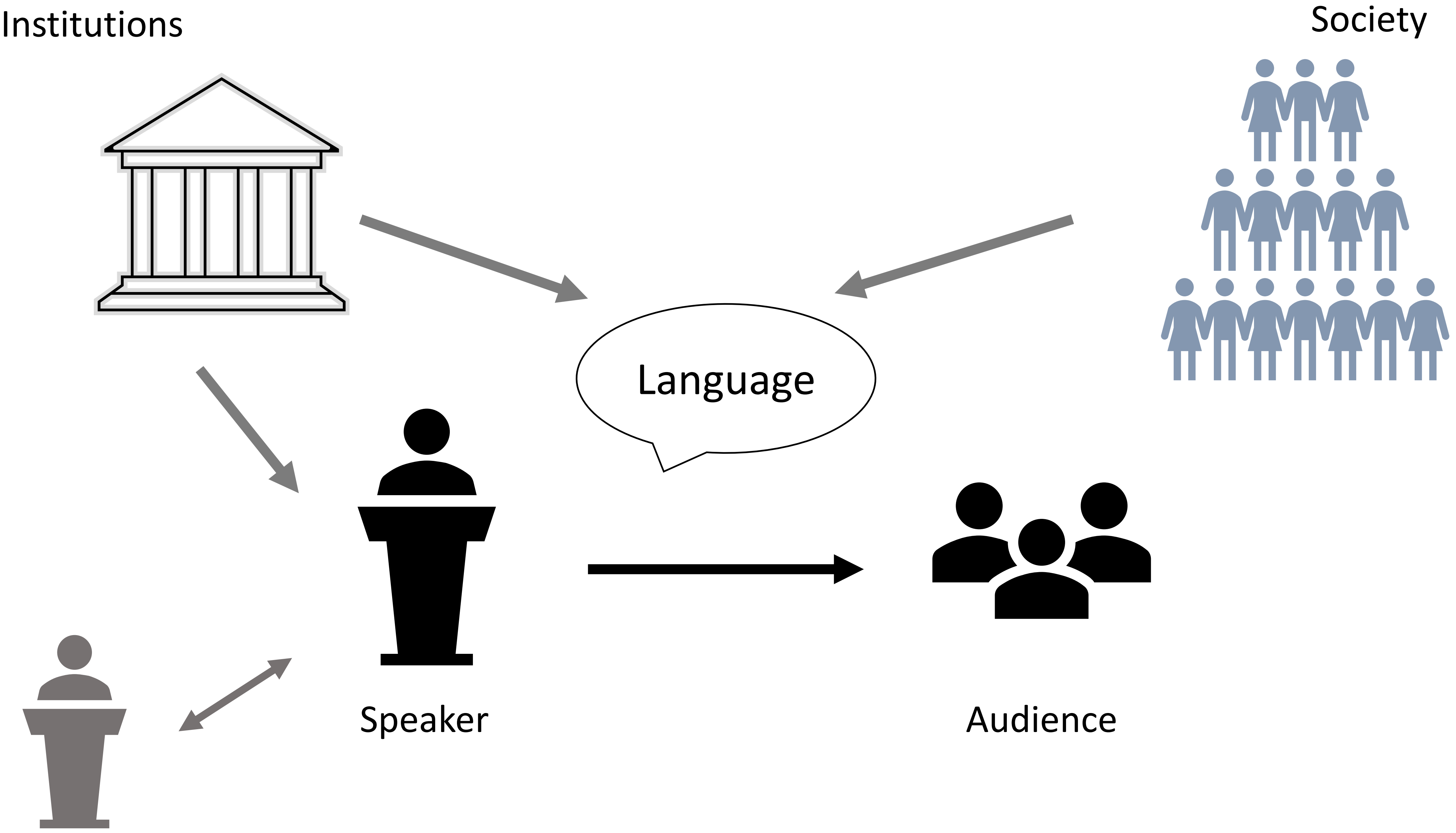
Validation

- 2 Reasons for Manual Coding
  - Training:
  - Validation

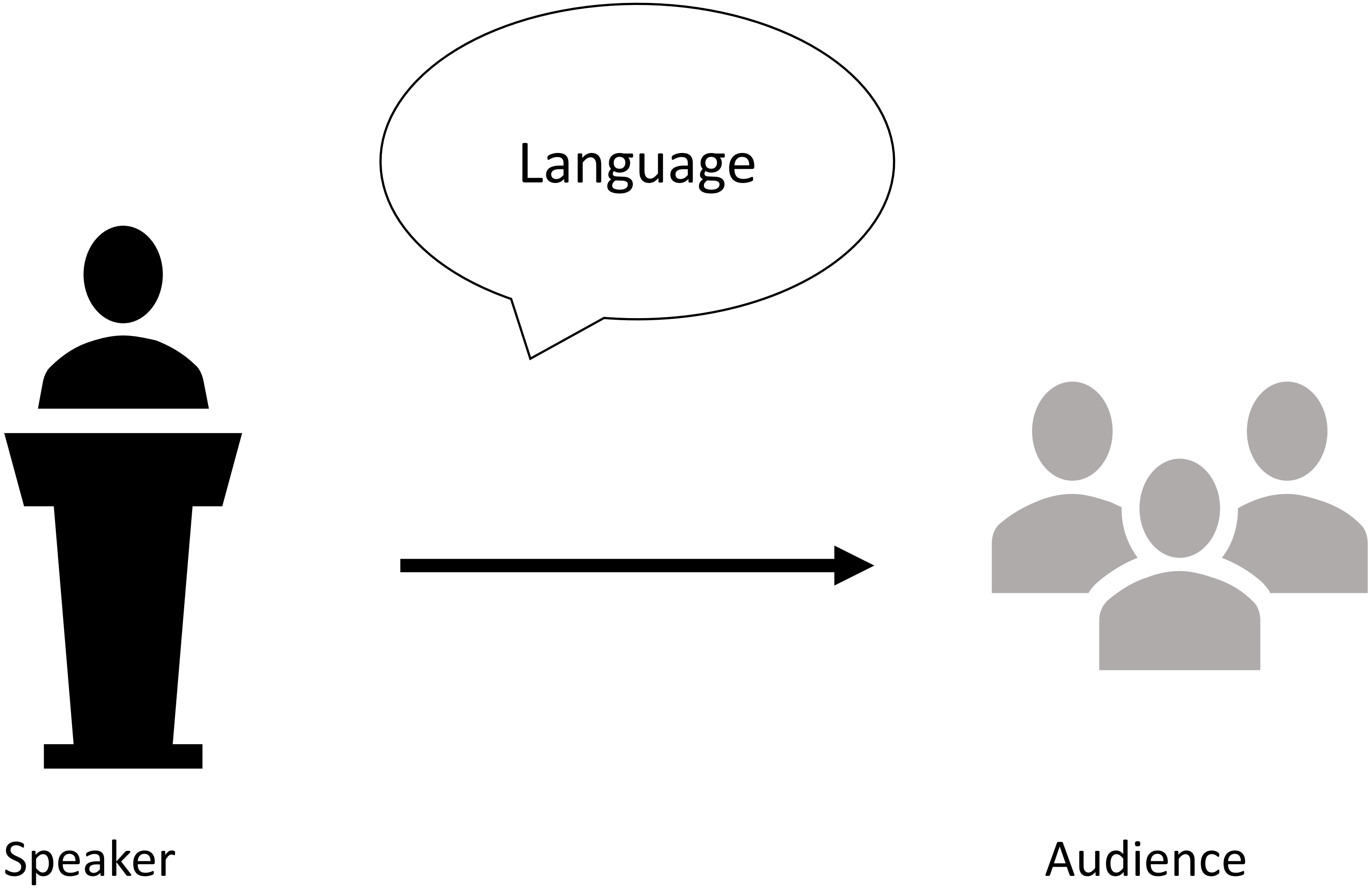
- Main differences:
  - Training requires a representation of every example
  - Validation requires a representation of examples necessary to succeed

- Validation is easier to code if random!

Taming Text

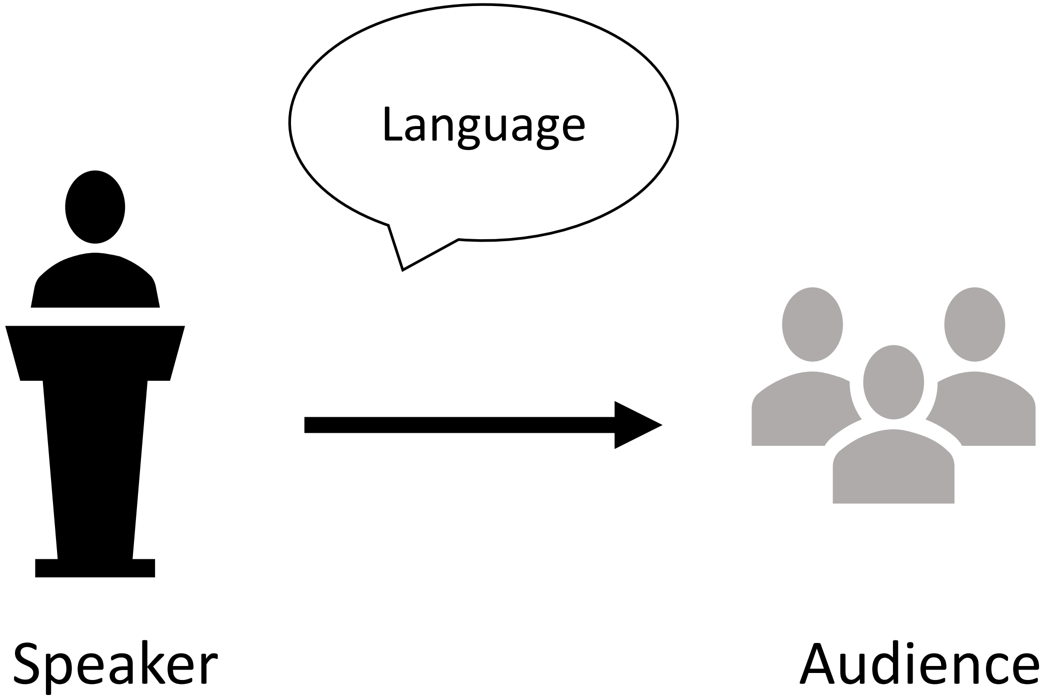

Institutions
Society
Language
Speaker
Audience

Speaker



Intent

- Inform
- Claim
- Critisize
- Define
- Defend
- Persuade



Language

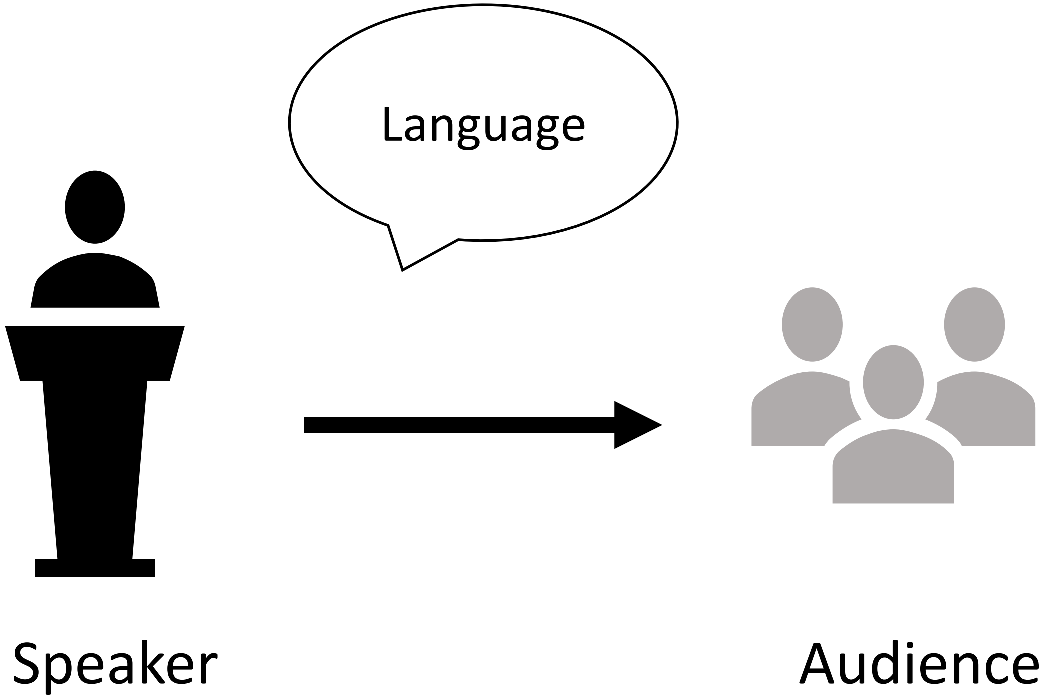
Speaker

Audience

Language as Action

Form

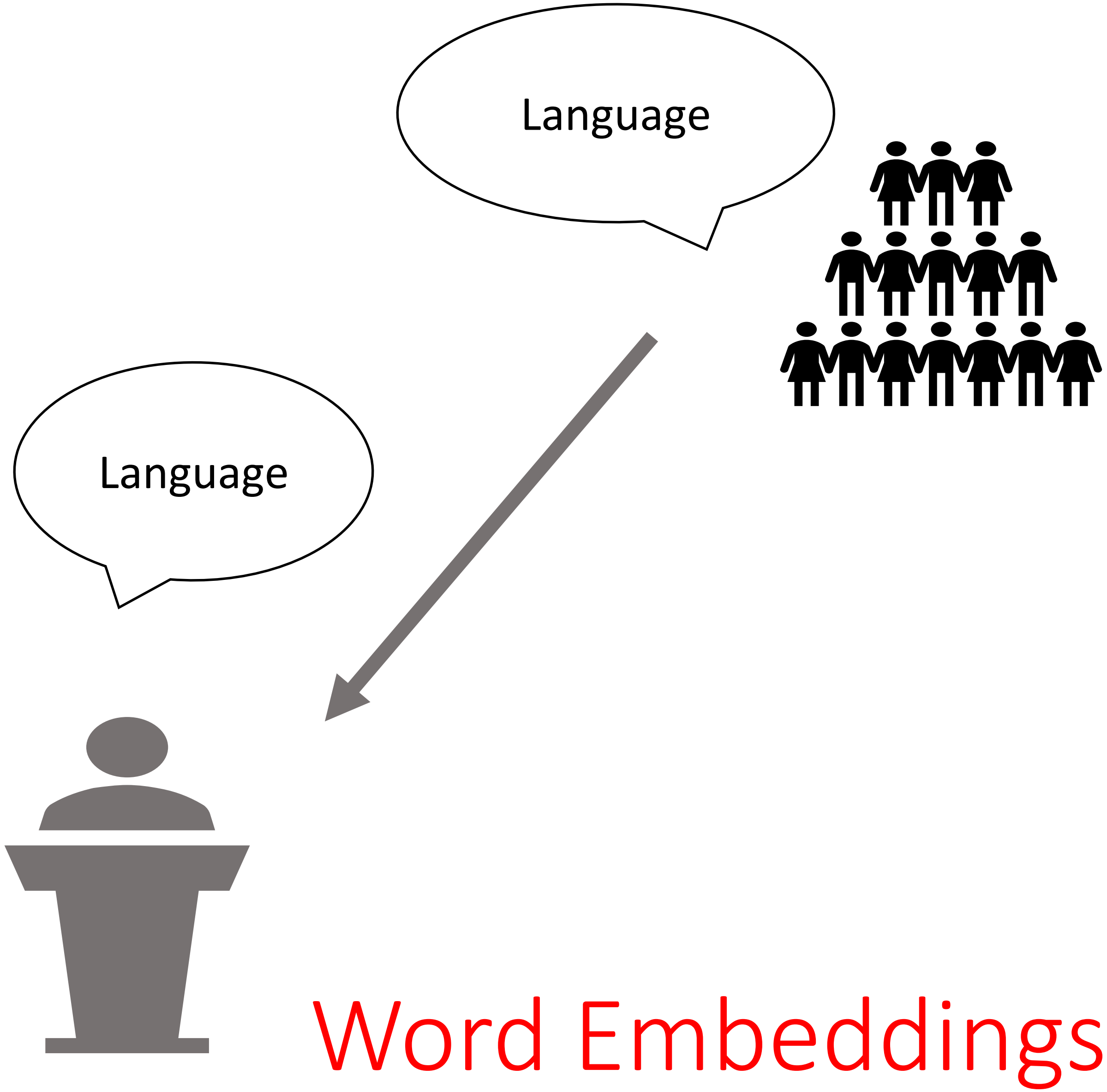
- Style
- Valence
  - Negativity
  - Stance
- Ambiguity
- Saliency
- Personlity



Language as Expression of Characteristics

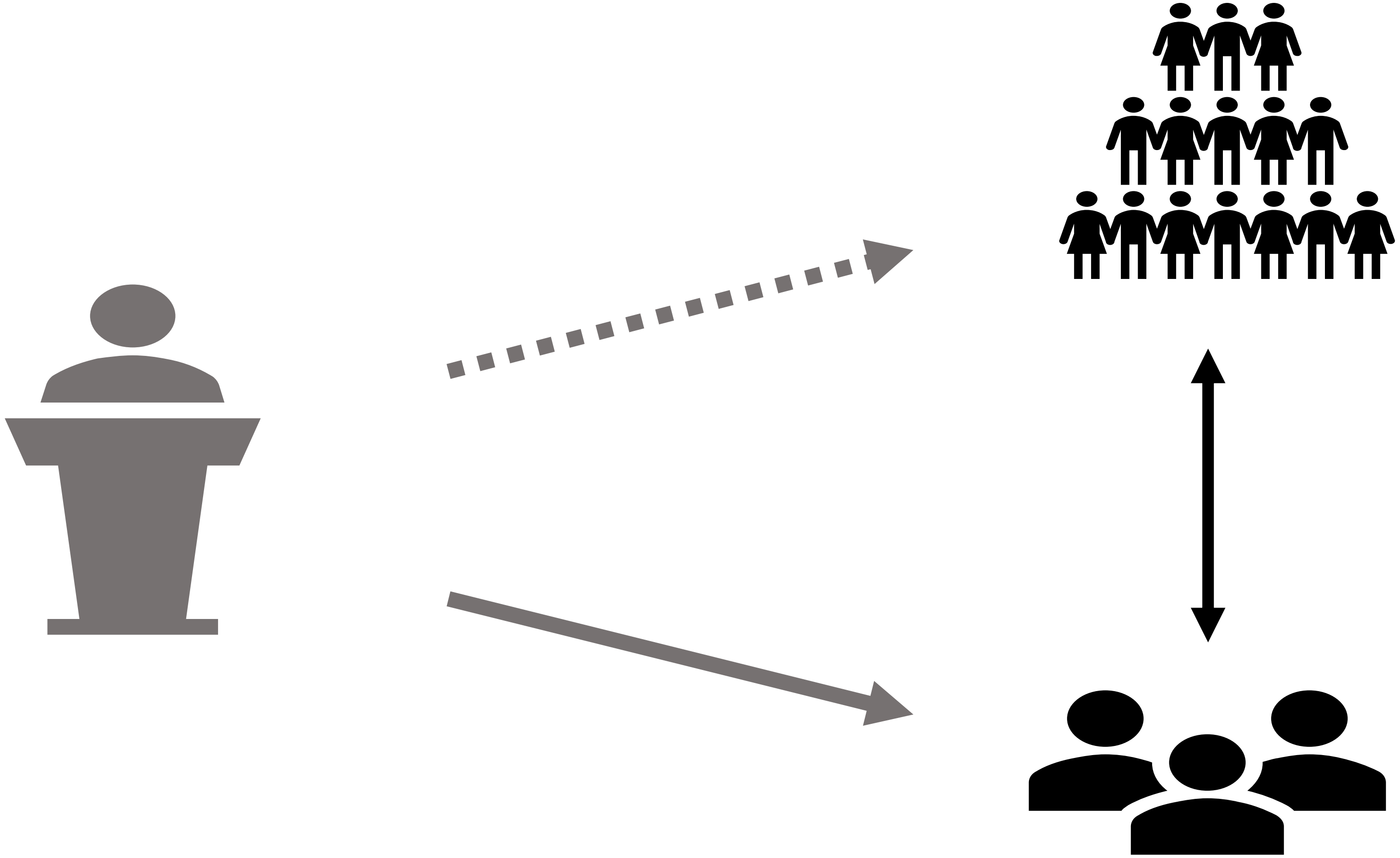
Language

- How is language used by society?
- How does it determine what can be said?



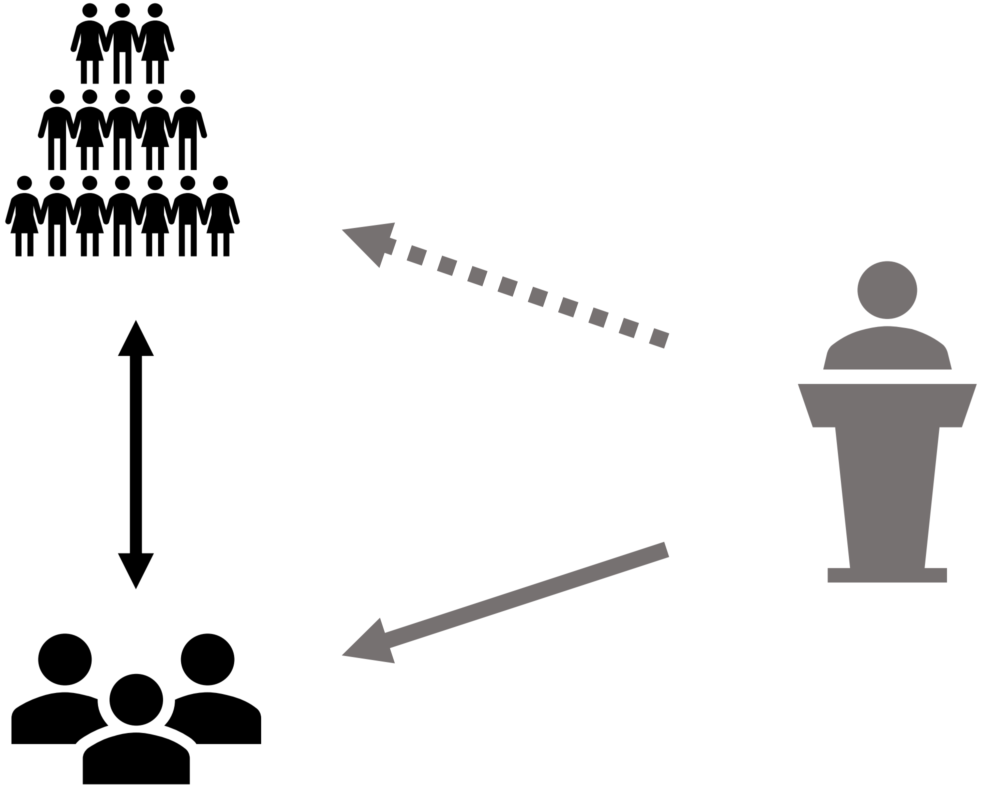
Word Embeddings?

Audience



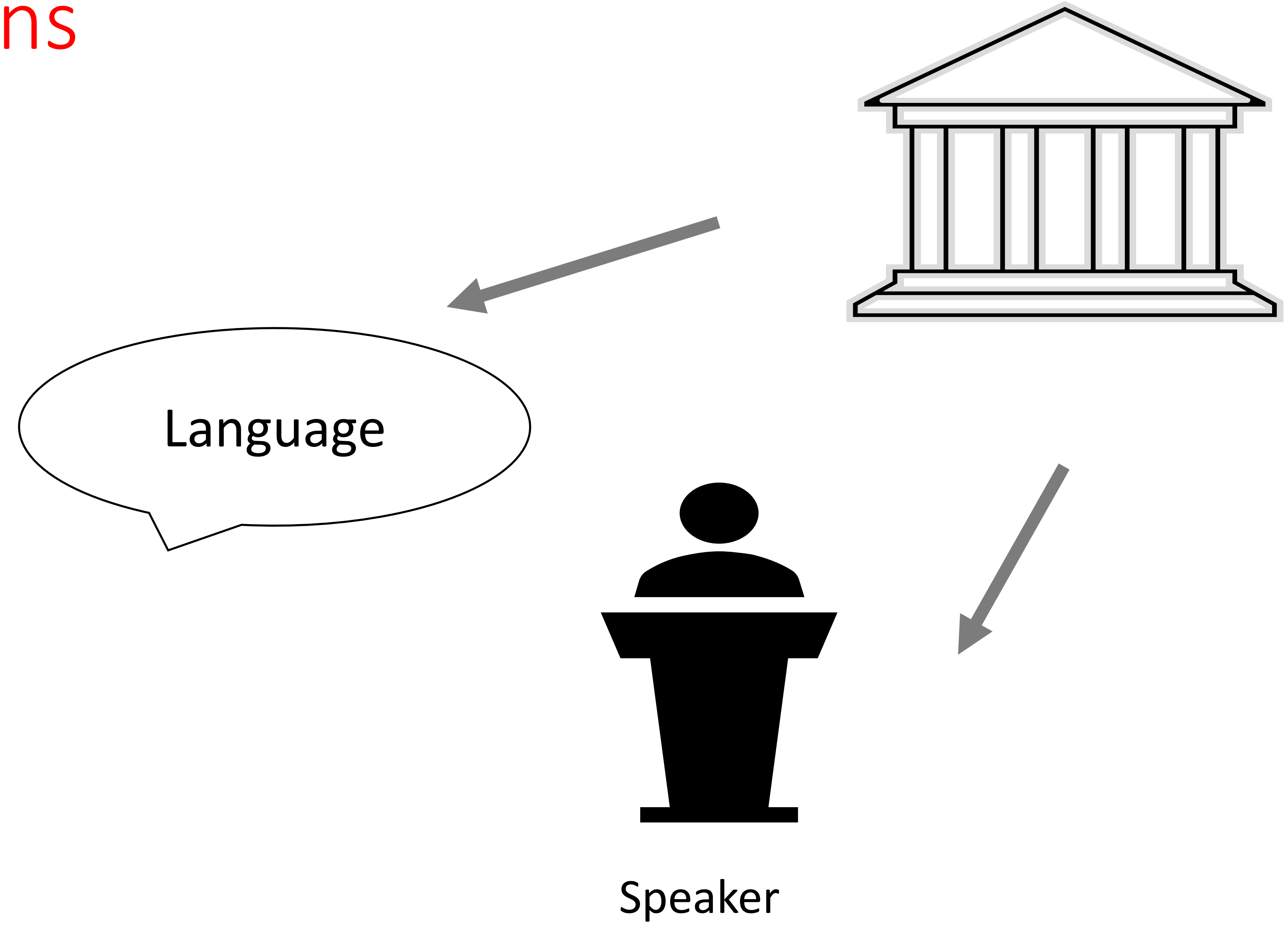
Audience

- Who is text addressed to?
- How will perception differ between audiences?
- Does audience shape message?



Local Embeddings!

Institutions

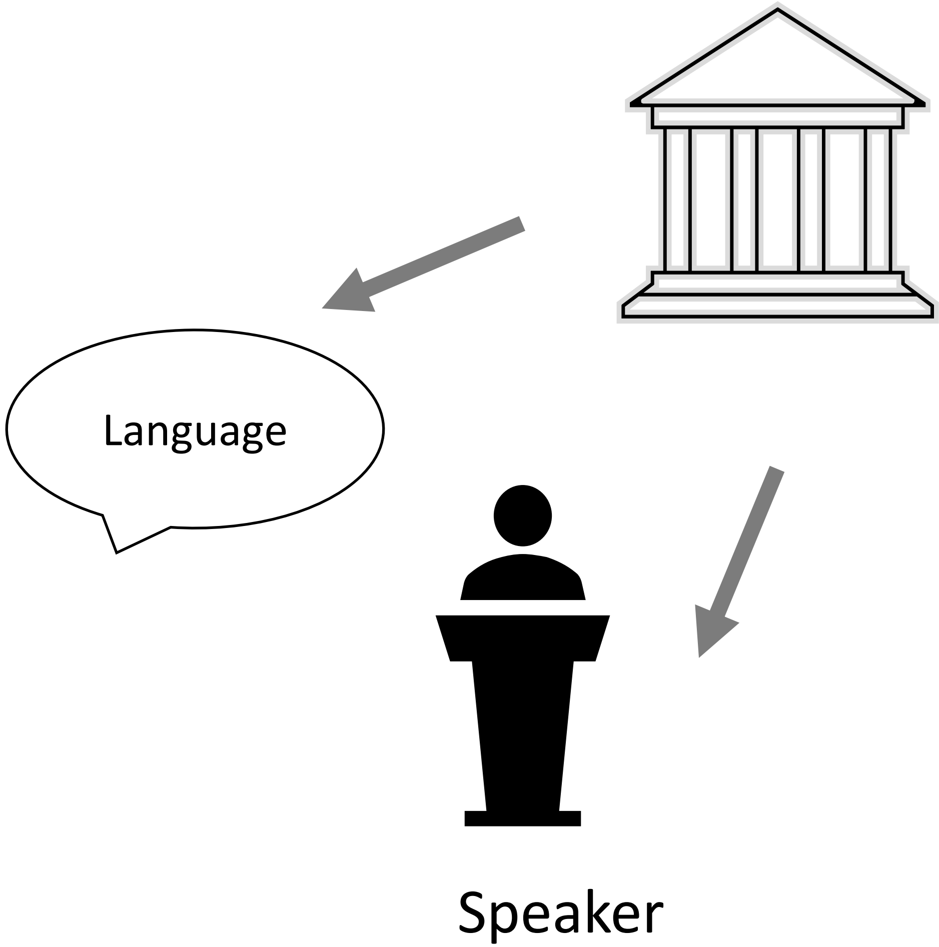


Language

Speaker

Institutions

- „Rules for Text“
  - Length
  - Structure
  - Social Norms
- Rules for Content
- Rules for Speakers



Meta Structures

- Level of Analysis
- Sentence
- Paragraph
- Document
- Meta Structure
  - Debate
  - Newspaper Issue
  - Category
- Corpus